



TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN, ĐHQG-HCM
ĐỀ THI KẾT THÚC HỌC PHẦN
Học kỳ I - Năm học 2021-2022

MÃ LƯU TRỮ

Tên học phần: Trực quan hóa dữ liệu

Mã HP:

Thời gian làm bài: 90 phút

Ngày thi: 16-01-2022

Ghi chú: Sinh viên **được phép** sử dụng tài liệu khi làm bài.

Câu 1 (1 điểm):

- Đọc file csv
- In ra 5 rows đầu tiên.
- In ra info và describe của file đã cho.
- Cho biết số lượng records và features.
- Cho biết các trường missing values.

Câu 2 (1 điểm):

- Cleaning lại cột Age, giá trị của cột Age nằm trong giới hạn từ 0 đến 100 ([0:100]), nếu không thuộc giới hạn trên thì thay bằng Nan (np.nan).
- Vẽ biểu đồ phân bố độ tuổi (Age Distrubtion) từ cột Age đã drop nan.

Câu 3 (1 điểm):

- Vẽ biểu đồ biểu diễn 10 quốc gia có số lượng records nhiều nhất ("Top 10 Country")

Câu 4 (1 điểm):

- Vẽ biểu đồ phân bố ở cột treatment (Treatment Distribution).

Câu 5 (1 điểm):

- Thống kê số lượng yes/no ở cột family_history theo các nhóm tuổi : <18, 18-24, 25-34, 35-44, 45+ ở cột Age.
Chú ý: Cột Age không lấy giá trị Nan.
- Từ đó vẽ biểu đồ biểu diễn số lượng yes/no của family_history theo từng nhóm tuổi.

Câu 6 (1 điểm):

- Tương tự câu 5, nhưng thay cột family_history thành cột treatment.

Câu 7 (1 điểm):

- Vẽ biểu đồ phân bố ở cột work_interfere (work_nterfere Distribution)
- Xác định các unique và số lượng tương ứng của cột no_employee và vẽ biểu đồ biểu diễn các unique này theo tỉ lệ phần trăm.
Chú ý: Nhớ thực hiện bỏ missing values.

Câu 8 (1 điểm):

- Vẽ biểu đồ biểu diễn số lượng yes/no của cột tech_company theo từng unique của cột no_employee
- Tương tự ý trên nhưng thay cột tech_company thành remote_work.

Câu 9 (1 điểm):

Hiện tại cột Gender có rất nhiều giá trị khác nhau ví dụ:

Nam : thì có giá trị Male, M, male ... Nữ: thì có giá trị F, female, Female.

- Chuẩn hóa cột Gender về 3 giá trị: Male, Female, Another (Nếu không xác định được).
- Vẽ biểu đồ biểu diễn số lượng Male, Female, Another trong cột Gender sau chuẩn hóa.

Câu 10 (1 điểm):

Dataset là khảo sát về sức khỏe tinh thần (mental health) và thông tin các cột như sau:

- Age : Respondent age
- Gender: Respondent gender
- treatment: Have you sought treatment for a mental health condition?
- family_history : Do you have a family history of mental illness?
- work_interfere: If you have a mental health condition, do you feel that it interferes with your work?
- no_employee: How many employees does your company or organization have?
- tech_company: Is your employer primarily a tech company/organization?
- remote_work: Do you work remotely (outside of an office) at least 50% of the time?

Và từ các biểu đồ bạn đã vẽ, bạn có nhận xét gì sức khỏe tinh thần (mental health) của dataset đã cho?